

Scotland's Rural College

## Development of a fixed list of terms for the Qualitative Behavioural Assessment of shelter dogs

Arena, Laura; Wemelsfelder, Françoise; Messori, Stefano; Ferri, Nicola; Barnard, Shanis

*Published in:*  
PLOS ONE

*DOI:*  
[10.1371/journal.pone.0212652](https://doi.org/10.1371/journal.pone.0212652)

First published: 04/10/2019

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication](#)

### *Citation for pulished version (APA):*

Arena, L., Wemelsfelder, F., Messori, S., Ferri, N., & Barnard, S. (2019). Development of a fixed list of terms for the Qualitative Behavioural Assessment of shelter dogs. *PLoS ONE*, 14(10), [e0212652].  
<https://doi.org/10.1371/journal.pone.0212652>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

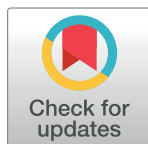
RESEARCH ARTICLE

# Development of a fixed list of terms for the Qualitative Behavioural Assessment of shelter dogs

Laura Arena<sup>1,2\*</sup>, Françoise Wemelsfelder<sup>3</sup>, Stefano Messori<sup>1</sup>, Nicola Ferri<sup>1</sup>, Shanis Barnard<sup>4</sup>

**1** Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', Teramo, Italy, **2** Università di Teramo, Facoltà di Medicina Veterinaria, Piano d'Accio, Teramo, Italy, **3** Animal & Veterinary Sciences, Scotland's Rural College, Easter Bush, Midlothian, Scotland, United Kingdom, **4** Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, West Lafayette, IN, United States of America

\* [laurarena@hotmail.it](mailto:laurarena@hotmail.it)



## Abstract

The shelter environment may have a severe impact on the dogs' quality of life, and there is thus a need to develop valid tools to assess their welfare. These tools should be sensitive not only to the animals' physical health but also to their mental health, including the assessment of positive and negative emotions. Qualitative Behaviour Assessment (QBA) is a 'whole animal' measure that captures the expressive quality of an animal's demeanour, using descriptive terms such as 'relaxed', 'anxious', and 'playful'. In this study, for the first time, we developed and tested a fixed-list of qualitative QBA terms for application to kennelled dogs. A list of 20 QBA terms was developed based on literature search and an expert opinion survey. Inter-observer reliability was investigated by asking 11 observers to use these terms to score 13 video clips of kennelled dogs. Principal Component Analysis (PCA) was used to extract four main dimensions explaining 70.9% of the total variation between clips. PC1 characterised curious/playful/excitable/sociable demeanour, PC2 ranged from comfortable/relaxed to anxious/nervous/stressed expression, PC3 described fearful demeanour, and PC4 characterised bored/depressed demeanour. Observers' agreement on the ranking of video clips on these four expressive dimensions was good (Kendall's W: 0.60–0.80). ANOVA showed a significant effect of observer on mean clip score on all PCs ( $p < 0.05$ ), due to few observers scoring differently from the rest of the group. Results indicate the potential of the proposed list of QBA terms for sheltered dogs to serve, in alignment with other measures, as a non-invasive assessment tool. However, the observer effect on mean PC scores points towards the need for adequate observer training, particularly in live scoring conditions. The QBA scoring tool can be integrated with existing welfare assessment protocols for shelter dogs and strengthen the power of those protocols to evaluate the animals' experience in shelters.

## OPEN ACCESS

**Citation:** Arena L, Wemelsfelder F, Messori S, Ferri N, Barnard S (2019) Development of a fixed list of terms for the Qualitative Behavioural Assessment of shelter dogs. PLoS ONE 14(10): e0212652. <https://doi.org/10.1371/journal.pone.0212652>

**Editor:** I. Anna S. Olsson, Universidade do Porto Instituto de Biologia Molecular e Celular, PORTUGAL

**Received:** January 8, 2019

**Accepted:** September 19, 2019

**Published:** October 4, 2019

**Copyright:** © 2019 Arena et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Thousands of dogs around the world are kept in temporary or permanent confinement for a variety of reasons and in a way that may affect their welfare [1,2]. There is evidence that shelter environments may have a severe impact on the quality of life of dogs [3,4]. This is likely due to factors such as social isolation and novel surroundings [5], especially if these are protracted over long periods of time [6]. For this reason, there has been an increasing interest by the scientific community to develop validated tools to assess the welfare of sheltered dogs [3,7].

Over the last decades, the concept of animal welfare has evolved from focusing primarily on the animal's physical health and ability to cope with its environment [8], to recognising that animals are sentient beings capable of experiencing positive and negative emotions [9]. It is now accepted that animals, though healthy, can nevertheless experience negative emotions when housed in unsuitable environments [10]. In addition the concept of animal welfare has shifted in recent times from focusing on the reduction of negative emotions (e.g. fear, pain) to ensuring that captive/domestic animals are also experiencing positive emotions (e.g. pleasure, happiness) [11,12]. It is therefore essential that animal welfare assessment tools include measures of positive welfare.

The potential of qualitative methods for assessing behaviour and welfare to play a role in such developments, and, alongside quantitative methods, to contribute to useful measures of positive animal welfare, has been a subject of review and discussion [13,14]. Qualitative Behavioural Assessment (QBA) is one such method which to date has generated a substantial body of research assessing emotional expressivity in a range of animal species [15–18]. QBA is a 'whole-animal' approach measuring not so much *what* the animal is doing physically, as *how* it is performing these behaviours, i.e. the expressive style or demeanour with which the animal is moving [19]. QBA uses a range of positive and negative qualitative terms such as 'relaxed', 'sociable', 'anxious' or 'fearful' in order to quantify an animal's experience in different situations and conditions [19].

The descriptive terms used in QBA can either be developed through an experimental procedure known as Free-Choice Profiling (FCP), in which each assessor generates his/her own terms based on the observation of animals in different situations [20,21], or are provided in the form of a pre-determined list of QBA terms given to observers to assess animals. When standardisation of measurement tools is required, for example for the purpose of on-farm welfare inspections, the use of a fixed-list designed for the purpose of welfare assessment in a particular species is more feasible than FCP. Pre-fixed QBA protocols are for example included in Welfare Quality<sup>®</sup> protocols for cattle, poultry and pigs [22] and in AWIN protocols for sheep, goats, horses, and donkeys [23–25], where they serve as a measure for positive emotional state.

Initial validation of the application of QBA to dogs was done by Walker et al. [26,27]. In these two studies Walker and colleagues applied FCP for the first time to assessing emotional expressivity in working dogs (all Beagles), and in shelter dogs in home pen and novel test pen environments. Outcomes showed, in the different dog groups, overlapping dimensions of emotional expressivity such as confidence/contentment, anxiousness/unsureness, alertness/attentiveness or curiousness/inquisitiveness. Walker et al. [27] also showed significant and meaningful correlations between QBA dimensions and quantitative ethogram-based behavioural measures, suggesting the potential value of QBA as a welfare assessment tool for dogs. Arena and colleagues [28] further investigated this potential by applying FCP to shelter dogs in a wide variety of shelter environments and social contexts including outdoor/indoor pens, single/pair/group housing, and the presence/absence of human activity. Results showed that by presenting more complex contexts and giving the animals more opportunities to express a wider repertoire of emotions, the observers generated 'richer' expressive dimensions than was

the case in Walker et al.'s [26,27] studies which used relatively standardised experimental settings.

The present study aimed to build on these outcomes by developing a fixed-term QBA terms list that might be integrated with existing on-shelter welfare assessment protocols. Adding QBA could provide information complementary to that provided by quantitative measures, extending a protocol's power to identify and detect emotional shifts in dogs across the positive and negative emotional spectrum [7,27,28]. Recognising that QBA relies on context-specific qualitative judgment of behavioural expression [19], terms included in a pre-fixed list should be representative of the large range of behavioural expressions that dogs could potentially show in variable kennel conditions. To achieve this goal, we generated a list of suitable terms based on the available dog behaviour and welfare literature, and then used an expert opinion survey to refine this list into a final set of 20 QBA terms. The inter-observer reliability of these terms was subsequently investigated by instructing eleven observers to score the emotional expressivity of dogs viewed in a sample of videos reflecting a wide range of shelter conditions.

## Material and methods

### Ethics statement

No special permission for use of dogs in such behavioural studies is required in Italy, since dogs were observed during their daily life and within their familiar shelter pens. When dogs were exposed to people, no physical interaction was required. We sent a formal request to carry out the study to all shelters managers and received permission from all of them.

All procedures were performed in full accordance with Italian legal regulations and the guidelines for the treatments of animals in behavioural research and teaching of the Association for the Study of Animal Behavior (ASAB).

No IRB approval was sought for observers participation as observers, but they provided informed signed consent to participate in the study and they were fully informed about the purpose and background of the study. Students ranged in age from 25 and 34 years.

### Selection of terms

**Literature review.** In a previous QBA study by this research group [28], 13 observers generated a number of terms to describe the emotional expression of shelter dogs from videos, using a Free Choice Profiling (FCP) methodology [15]. The FCP in Arena et al. [28] generated over 50 terms and the analysis extracted three main consensus dimensions. From that work, we selected 16 high-loading terms characterising each of the three main dimensions (Table 1). These 16 terms were used as a starting-point for the current study. Additional qualitative terms were selected from the relevant scientific literature on dogs' emotions [26,27,29], behaviour [30], personality and temperament [31–34], and from other QBA research [13,35,36]. Thus, nine new terms were added (affectionate, aloof, attention-seeking, boisterous, excited, explorative, happy, self-confident and serene), creating a preliminary list of 25 terms.

**Table 1. List of terms summarising the highest positive and negative loadings on each consensus dimension generated through Free Choice Profile methodology in Arena et al. [28].**

	Positive	Negative
Dimension 1	playful/sociable/curious	bored/uncomfortable/apathetic
Dimension 2	relaxed/tranquil	nervous/alert/fearful
Dimension 3	stressed/bored/anxious	wary/timorous/hesitant

<https://doi.org/10.1371/journal.pone.0212652.t001>

**Expert opinion.** To check for the appropriateness, relevance and ease of understanding of the 25 terms on the preliminary list, we performed an expert opinion survey. For this purpose, we selected a panel of 15 international experts in the field of dog personality and behaviour, shelter dog welfare and QBA methodology. The panel was composed of 12 females and 3 males. A one-round on-line survey was arranged using the SurveyMonkey<sup>®</sup> platform. In an introductory letter, experts were told that the goal of the project was to develop a comprehensive list of terms for qualitative assessment of the emotional state and welfare of dogs housed in different shelter environments. Eight out of 15 experts answered the survey anonymously.

For each of the 25 terms on the list, the authors provided a brief semantic characterisation, similarly as done previously for donkeys, horses and goats [25,37,38]. Such characterisations cannot be regarded as sharply delineated definitions similar to those provided for conventional ethograms, but rather are brief depictions containing both qualitative and quantitative elements that together should illustrate the meaning of a term, as appropriate for particular species and contexts. In the present study the Oxford English Dictionary (online version) was used as a starting-point, and terms were then further adapted using information from previous QBA studies on dogs [26–28], and previous QBA studies using fixed lists [25,37,38]. The experts were asked to score each term on the basis of four brief statements using a likert-scale from 1 to 5, where 1 corresponded to ‘not at all’, and 5 to ‘completely’ (Table 2). Experts also had the possibility to add comments in a free space if wanted. Furthermore, at the end of the survey the experts could suggest a maximum of three terms that, in their opinion, were missing but that they considered important to describe the emotional state of sheltered dogs. This expert opinion was used to refine our original list by deleting and/or adding terms, in order to eliminate synonyms, select the terms most relevant to describing shelter dog emotional state, and ensure that terms were easy to understand.

**Generating the final list.** The SurveyMonkey<sup>®</sup> software automatically generated a matrix with the average scores assigned by the experts to each term. We then calculated the mean scores of each of the four statements for all terms (Table 2). To decide whether to keep or delete a term, we used those total mean scores to establish thresholds for each statement: to be considered for inclusion, a term had to receive a mean score of 3 or more in statements #1, #3 and #4, and less than 3 in statement #2. Furthermore, we summarised and classified the negative comments provided by the experts as ‘prone to anthropomorphism’, ‘too generic’, ‘too similar to other terms’, ‘is not an emotional state’. We put aside inapplicable comments such as those repeating the same concept expressed in the statement (e.g. “not useful to assess animal welfare”), those referring to the definition given by the authors (e.g. “I would include ‘interest’ in the definition”) and comments without justification (e.g. “I suggest to delete this term from the list”).

Finally, we defined inclusion/exclusion criteria in order to refine our list.

As exclusion criteria, we established that a term would be excluded from the list if it had at least:

**Table 2. Mean scores for the four statements presented in the survey for each term.**

In your opinion, this term is:	Survey mean scores <sup>a</sup>
1. easily comprehensible and intuitive	3.60
2. ambiguous, might be interpreted with different meanings	2.75
3. useful to describe the emotional state of sheltered dogs	3.49
4. relevant to assess the welfare of sheltered dogs	3.25

<sup>a</sup>calculated by averaging for each statement the experts’ scores across all terms.

<https://doi.org/10.1371/journal.pone.0212652.t002>

- two insufficient scores
- one insufficient score and one negative comment
- no insufficient scores, but three or more negative comments.

### Inter-observer reliability

The final list of 20 terms resulting from the previous steps was subsequently used to test for inter-observer reliability.

**Animals and video-clips.** A convenience sample of dogs in 13 kennels were video-recorded in seven different Italian shelters: three in Northern Italy (Emilia-Romagna and Piemonte Regions), two in the Centre (Abruzzo Region) and two in the South (Puglia Region). The dogs had been housed in the kennels on a long-term basis (> 6 months since admittance), either singly, in pairs, or in groups of 3 dogs or more. In Italy, group housing is a common practice. The aim was to record a sample of these dogs in different scenarios to capture a variety of dog behavioural expressions. The dogs were video-recorded in their kennels, with sound, using a smartphone (Samsung GT-I9100P) that had been mounted on a tripod positioned a few meters from the fence. Thus, a set of 13 video clips was obtained, with 4 clips showing singly housed dogs, 4 clips showing dogs housed in pairs, and 5 clips showing dogs in groups of 3 or more.

To generate a variety of behavioural expressions in the dogs, three scenarios (common in a shelter environment) were created, and the dog/s in each kennel was/were recorded for 2 minutes during one of these scenarios: 1. under normal conditions with no external intervention, 2. in the presence of an unknown person, or 3. in the presence of a familiar person. The unfamiliar person was one of two researchers (one female and one male), while the familiar person was a shelter operator available at the time of recording. The unfamiliar person followed a simple protocol, approaching the fence of the kennel and standing one metre from the fence ignoring the dog/s for 1 minute, and subsequently talking gently to the dog/s moving a hand slowly along the fence for another minute. Shelter operators were asked to enter the kennel and interact with the dog/s for 2 minutes. For the scenario of normal conditions, the dog/s were filmed for 2 minutes without any external intervention. In the two intervention scenarios the persons were visible in the videos. Scenarios were assigned to kennels so as to obtain a balanced distribution of scenarios across kennels and dog groupings.

The 2 minute clips thus obtained were edited (using the Avidemux 2.6.8 programme) so as to ensure that all dogs in a kennel (whether housed singly, in pairs or in groups) were visible throughout a clip, resulting in a set of 13 clips with an average length of 1.5 minutes. This final set of clips consisted of 5 clips in normal conditions with no external intervention (2 single dogs, 1 pair, and 2 groups), 4 clips with an unknown person (1 single, 2 pairs, and 1 group), and 4 clips with a familiar person (1 single, 1 pair, and 2 groups).

**Observers and observation session.** The observation session was carried out in Italy. Eleven participants to a course for dog trainers were recruited as observers (8 females and 3 males). All observers had previous experience with dogs and owned at least one dog. Five of them were dog trainers, four were volunteers at dog shelters, two had graduated at the Faculty of Animal Welfare and Protection, and one was a veterinarian. None of the observers had previous experience with QBA.

Before starting the assessment of the video clips, approximately 1 hour was dedicated to an introduction with the goal of explaining the aim of the study and the operative procedures. Observers were given a list of the QBA terms to be used for scoring with a brief characterisation of the meaning of each term (see paragraph above for a description of how these



characterisations were generated). These characterisations were translated onto Italian to facilitate scoring by the observers. Observers had the opportunity to discuss and align their understanding of these terms.

Observers were told that the aim of the study was to investigate whether observers can agree in using a fixed list of QBA terms to assess emotional expressivity in shelter dogs. Emotional expressivity was defined as an animal's dynamic style of interaction with its environment, conspecifics and humans (*how* the animal behaves as opposed to *what* it does). Observers were told that when scoring a pair or group of dogs housed together in a kennel, this was to be done not by assessing individual dogs and then adding them up, but by assessing the overall level of intensity at which a certain expressive quality was present in the interactions of the whole group. Each observer was provided with scoring sheets (one for each video clip) on which Visual Analogue Scales (VAS) of 125 mm of length were placed next to each term. Participants were told they must score the dog/s in each clip on every qualitative term on the list. They were instructed on how to use the VAS for scoring: the left end of the scale corresponded to the minimum score (0 mm), meaning the expressive quality indicated by the term was entirely absent in that dog or group of dogs, whereas the right end represented the maximum score (125 mm), meaning that the expressive quality indicated by the term was strongly dominant in that dog or group of dogs. Observers were asked to avoid talking during the session.

Once instructions had ended, observers watched the 13 clips projected onto a lecture hall screen and, after each clip, had approximately 2 mins to score the dogs' expressions on the rating scales, by drawing a vertical line across the VAS at the point they felt was appropriate. A score was assigned to each term for each clip by measuring with a ruler the distance in millimetres between the minimum point of the VAS and the point where the observer marked the line.

**Statistical analysis.** IBM SPSS Statistics 21 software (IBM Corp., 2012) was used for statistical analysis. The QBA scores gathered by the 11 observers for the 13 video clips on 20 QBA terms were all analysed together as part of one Principal Component Analysis (PCA, correlation matrix, varimax rotation). Before running the PCA, we explored the sampling adequacy by checking Kaiser-Meyer-Olkin (KMO) and anti-correlation image matrix values. PCA is a form of multi-variate analysis that can be used to reduce a large set of possibly correlated variables (QBA terms) to a small set of linearly uncorrelated variables (principle components or PCs, also referred to as dimensions), that still reflects most of the information contained in the large set. In other words, PCA serves to identify patterns of difference (dimensions) that explain most of the variation between units (video clips) as measured by a large set of variables (terms). The PCA assigned scores to each video clip for each of the 11 observers on each of the main extracted PCs/dimensions, and these scores were used as input for the analysis of inter-observer reliability in 2 different ways. First, how well the ranking of video clip scores from low to high on each of the main PCs was aligned between different observers was determined using Kendall's W coefficient of concordance [37]. Kendall's W values can vary from 0 (no agreement at all) to 1 (complete agreement), with values higher than 0.6 showing substantial agreement. Secondly, whether observers differed in the quantitative range occupied by clip scores on each PC, and the mean value resulting from that range, was investigated using a one-way ANOVA with observer as fixed factor and video clip as random factor. Post-hoc multiple comparisons (using Tukey HSD test) and homogeneous subsets were inspected to check which and how many observers were in disagreement. Finally, we also assessed alignment between observers in their ranking of clips on each separate QBA term, using the Kendall's W test.

## Results

### Generation of the list of terms

Mean scores attributed to each statement and for each term by the panel of responding experts are summarised in Table 3.

On the basis of the inclusion/exclusion criteria, nine terms were deleted from the list: boisterous, aloof, timorous, tranquil, serene, apathetic (all had at least two insufficient scores, and

**Table 3. Mean scores attributed by the experts to each term for each statement.** #Q: key words “comprehensible”, “ambiguous”, “useful” and “relevant” correspond to the statements as listed in Table 2.

Terms	#Q	Score	Terms	#Q	Score	Terms	#Q	Score
Playful	Comprehensible	4.38	Nervous	Comprehensible	3.63	Wary	Comprehensible	3.75
	Ambiguous	1.75		Ambiguous	2.75		Ambiguous	3.00
	Useful	4.13		Useful	3.75		Useful	4.13
	Relevant	4.13		Relevant	3.13		Relevant	3.88
Sociable	Comprehensible	4.25	Alert	Comprehensible	3.75	Hesitant	Comprehensible	3.88
	Ambiguous	1.88		Ambiguous	2.75		Ambiguous	2.75
	Useful	4.00		Useful	4.00		Useful	3.00
	Relevant	4.13		Relevant	3.00		Relevant	2.75 <sup>a</sup>
Curious	Comprehensible	3.75	Boisterous <sup>b</sup>	Comprehensible	2.38 <sup>a</sup>	Aloof <sup>b</sup>	Comprehensible	3.00
	Ambiguous	2.00		Ambiguous	3.75 <sup>a</sup>		Ambiguous	3.13 <sup>a</sup>
	Useful	3.50		Useful	2.88 <sup>a</sup>		Useful	2.63 <sup>a</sup>
	Relevant	3.25		Relevant	2.50 <sup>a</sup>		Relevant	2.75 <sup>a</sup>
Happy <sup>b</sup>	Comprehensible	3.88	Excited	Comprehensible	3.25	Fearful	Comprehensible	4.75
	Ambiguous	2.75		Ambiguous	2.63		Ambiguous	1.63
	Useful	3.50		Useful	3.75		Useful	4.50
	Relevant	3.25		Relevant	3.00		Relevant	4.25
Affectionate <sup>b</sup>	Comprehensible	3.25	Apathetic <sup>b</sup>	Comprehensible	3.13	Timorous <sup>b</sup>	Comprehensible	3.25
	Ambiguous	2.13		Ambiguous	3.13 <sup>a</sup>		Ambiguous	3.25 <sup>a</sup>
	Useful	3.38		Useful	2.88 <sup>a</sup>		Useful	3.38
	Relevant	2.88 <sup>a</sup>		Relevant	2.63 <sup>a</sup>		Relevant	2.88 <sup>a</sup>
Attention-seeking	Comprehensible	3.88	Uncomfortable	Comprehensible	3.25	Stressed	Comprehensible	3.88
	Ambiguous	2.38		Ambiguous	2.75		Ambiguous	2.38
	Useful	3.38		Useful	3.25		Useful	4.38
	Relevant	3.25		Relevant	3.50		Relevant	4.63
Relaxed	Comprehensible	3.50	Bored	Comprehensible	3.25	Anxious	Comprehensible	4.25
	Ambiguous	2.88		Ambiguous	2.75		Ambiguous	1.88
	Useful	3.50		Useful	3.50		Useful	4.50
	Relevant	3.25		Relevant	3.50		Relevant	4.50
Tranquil <sup>b</sup>	Comprehensible	2.50 <sup>a</sup>	Self-confident <sup>b</sup>	Comprehensible	3.50	Explorative	Comprehensible	3.88
	Ambiguous	3.25 <sup>a</sup>		Ambiguous	2.88		Ambiguous	2.00
	Useful	2.88 <sup>a</sup>		Useful	3.00		Useful	3.25
	Relevant	2.63 <sup>a</sup>		Relevant	2.88 <sup>a</sup>		Relevant	3.38
Serene <sup>b</sup>	Comprehensible	2.88 <sup>a</sup>						
	Ambiguous	3.50 <sup>a</sup>						
	Useful	2.00 <sup>a</sup>						
	Relevant	2.00 <sup>a</sup>						

<sup>a</sup> ‘Insufficient’ score

<sup>b</sup> Deleted terms.

<https://doi.org/10.1371/journal.pone.0212652.t003>



some also had negative comments), affectionate, self-confident (one insufficient score and at least one negative comment), happy (at least three negative comments). We evaluated the additional terms suggested by the panel of experts, and we added interested, depressed and aggressive (each one, suggested three times) and reactive (suggested twice). Finally, the term uncomfortable was replaced by its positive form 'comfortable', because this avoids having to score a double negative (i.e. lower scores implying an animal to be not at all uncomfortable) and poses less risk of confusion for observers.

As a result, the final list was composed of 20 terms (Table 4). Overall this list was balanced in term of their semantic meaning (*positive* versus *negative* emotional state) and level of arousal they expressed (*high* versus *low* arousal). As before, brief characterisations of each QBA term were produced using procedures described in the methods above.

### Inter-observer reliability of terms

KMO was well above the threshold of 0.50 (0.83) and the anti-image correlations were also high for all individual variables ( $> 0.55$ ). After visual inspection of the scree plot, four components were extracted which together explained 70.9% of the total variance and all with Eigenvalues  $> 1$  (Table 5). The observers' agreement in the ranking of video clips on each of these four dimensions ranged from 0.60 to 0.80, all significant at  $p < 0.001$  (Table 5).

Table 6 shows the loadings of each term on the four principal components (PC). PC1 was characterized by positive terms curious/attention-seeking/playful/excited/sociable/interested and explorative; PC2 characterized dogs as ranging from comfortable/relaxed to anxious/nervous/stressed. Fig 1 shows the 20 QBA descriptive terms plotted along these first two PCs. PC3 was characterized by the terms fearful/hesitant/wary, and finally, PC4 by the terms depressed/bored. Fig 2 shows the 20 QBA terms plotted along PC3 and PC4.

**Table 4. Final list of terms and their characterisations.**

Term	Definition
aggressive	impetuous, shows signs and posture of defensive or offensive aggression
alert	vigilant, inquisitive, on guard
anxious	worried, unable to settle or cope with its environment, apprehensive
attention-seeking	interactive, looking for contact/interaction, vying for people's attention, affectionate
bored	disinterested, passive, showing sub-optimal arousal levels/drowsiness signs
comfortable	without worries, settled in its environment, peaceful with other dogs, people and external stimuli
curious	actively interested in people or things, explorative, inquiring, in a positive relaxed manner
depressed	dull, sad demeanour, disengaged from and unresponsive to the environment, quiet, apathetic
excited	positively agitated in response to external stimuli, euphoric, exuberant, thrilled
explorative	confident in exploring the environment or new stimuli, investigative
fearful	timid, scared, timorous, doesn't approach people or moves away, shows postures typical of fear
hesitant	unsure, doubtful, shows conflicting behaviour, uncertain whether to approach or trust a stimulus, other dog or person
interested	attentive, attracted to stimuli and attempting to approach them
nervous	uneasy, agitated, shows fast arousal, unsettled, restless, hyperactive
playful	cheerful, high spirits, fun, showing play-related behaviour, inviting others to play
reactive	responsive to external stimuli
relaxed	easy going, calm or acting in a calm way, doesn't show tension
sociable	confident, friendly toward humans and other dogs, appreciates human attentions, shows greeting behaviour
stressed	tense, shows signs of distress
wary	cautious, prudent, suspicious, circumspect

<https://doi.org/10.1371/journal.pone.0212652.t004>

Table 5. Principal component (PC) analysis outcomes and inter-observer agreement (using Kendall's W) for QBA rating scales.

	PC1	PC2	PC3	PC4
Eigenvalue	5.66	5.18	1.78	1.56
% explained variance	28.3	25.9	8.9	7.8
% cumulative variance	28.3	54.2	63.1	70.9
Kendall W	0.61	0.80	0.71	0.60
p-value	p<0.001	p<0.001	p<0.001	p<0.001

<https://doi.org/10.1371/journal.pone.0212652.t005>

ANOVA showed that there was a significant effect of observer on mean clip scores on all 4 PCs (PC1:  $F = 13.7$ ,  $p < 0.001$ ; PC2:  $F = 3.5$ ,  $p < 0.001$ ; PC3:  $F = 15.6$ ,  $p < 0.001$ ; PC4:  $F = 6.4$ ,  $p < 0.001$ ). This means that although observers were in agreement on the ranking of clips on the 4 PCs (see above), they differed with respect to the mean value generated by PCA for clip scores on the 4 PCs. Inspection of post-hoc multiple comparisons indicated that for PCs 1, 3 and 4 most observers ( $>7$ ) presented homogeneous means, with the remaining 3–4 observers scoring differently from the rest of the group. The observers scoring differently were often the same across all PCs. Post-hoc analysis for PC2 highlighted one clear outlier (observer 10), with the remaining observers not differing from each other in mean clip score. When running the ANOVA again, without observer 10, the observer effect on PC2 was not significant ( $F = 1.62$ ,  $p = 0.12$ ).

The Kendall W values for separate QBA terms were all significantly different from chance ( $p < 0.001$ ), but only 8 terms reached values higher than 0.60, while the term 'depressed' showed a particularly low value (Table 7).

Table 6. Loadings for each QBA term on the four principal components (PC1, PC2, PC3 and PC4) extracted by the PCA analysis.

Variable	PC1	PC2	PC3	PC4
curious	0.885 <sup>a</sup>	-0.020	-0.095	-0.168
attention-seeking	0.839 <sup>a</sup>	0.090	-0.236	0.102
playful	0.834 <sup>a</sup>	-0.057	-0.145	0.016
excited	0.809 <sup>a</sup>	0.108	0.013	-0.132
sociable	0.779 <sup>a</sup>	-0.124	-0.364	0.074
interested	0.770 <sup>a</sup>	-0.010	0.211	-0.230
explorative	0.627 <sup>a</sup>	-0.116	0.420	-0.180
anxious	0.102	0.870 <sup>a</sup>	0.193	0.072
nervous	0.104	0.864 <sup>a</sup>	0.164	-0.075
stressed	-0.046	0.843 <sup>a</sup>	0.199	0.156
comfortable	0.245	-0.765 <sup>a</sup>	-0.217	0.231
relaxed	0.046	-0.761 <sup>a</sup>	-0.172	0.245
reactive	0.511	0.514	-0.030	-0.406
fearful	-0.068	0.299	0.818 <sup>a</sup>	0.026
hesitant	-0.039	0.104	0.803 <sup>a</sup>	0.157
wary	-0.180	0.352	0.777 <sup>a</sup>	0.001
alert	0.035	0.499	0.596	-0.318
depressed	-0.181	0.130	0.161	0.807 <sup>a</sup>
bored	-0.201	-0.276	0.001	0.758 <sup>a</sup>
aggressive	-0.101	0.144	0.339	-0.385

<sup>a</sup> Loadings higher than 0.60

<https://doi.org/10.1371/journal.pone.0212652.t006>

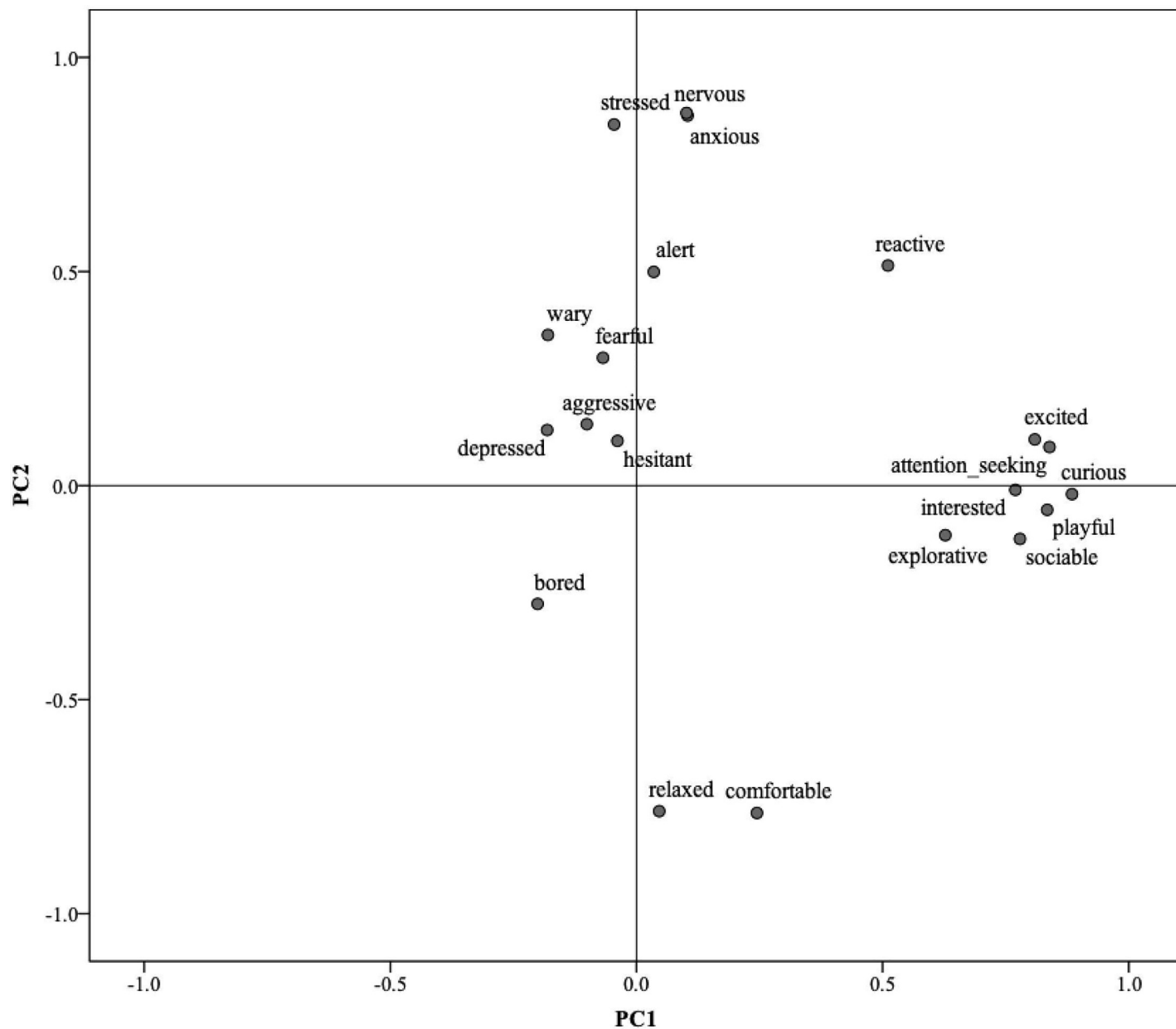
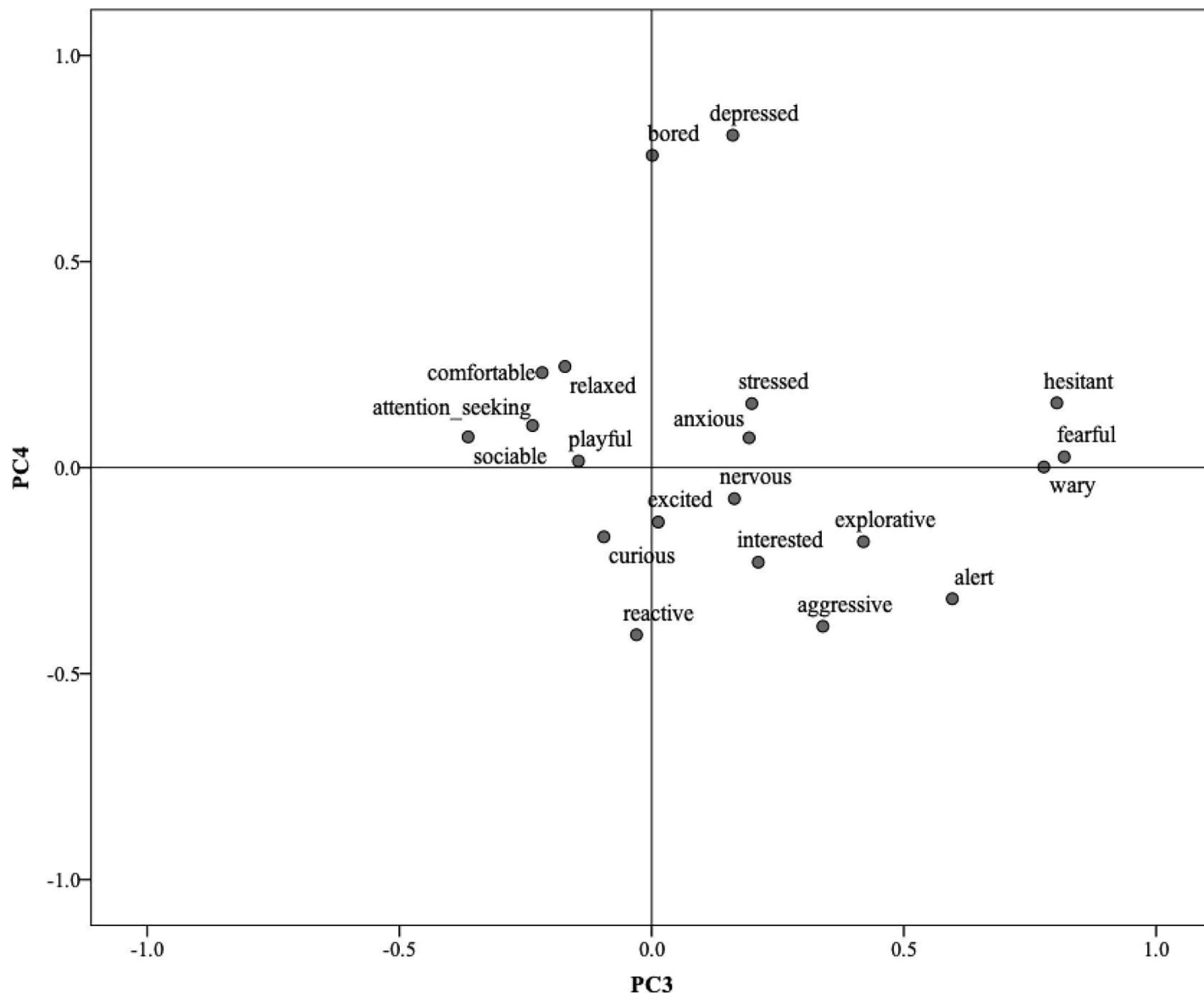


Fig 1. Distribution of the 20 QBA terms on Principal Component 1 (PC1) and Principal Component 2 (PC2).

<https://doi.org/10.1371/journal.pone.0212652.g001>

## Discussion

The aim of this research was to develop and test the inter-observer reliability of a fixed Qualitative Behaviour Assessment (QBA) rating scale for the purpose of field assessment of emotional expressivity in dogs in a shelter environment. Attempts had previously been made to develop and validate welfare assessment protocols for kennelled dogs [4,7], however, these protocols did not include measures addressing animals' emotions. To our knowledge, our study is the first to generate a fixed QBA rating scale for the assessment of shelter dogs' emotion. A previous Free Choice Profiling study served as a basis for generating an initial list of terms [28], and subsequently we searched the literature and consulted expert opinion to create a final list of 20



**Fig 2. Distribution of the 20 QBA terms on Principal Component 3 (PC3) and Principal Component 4 (PC4).**

<https://doi.org/10.1371/journal.pone.0212652.g002>

terms covering as much as possible the relevant behavioural expressions appropriate for the assessment of dogs in kennels. Eleven observers then used this list to score kennelled dogs in 13 video clips, and PCA of these scores generated 4 Principal Components (PCs) explaining 70.9% of the total variance. PC1 was characterised by the terms 'curious/attention-seeking/playful/excited/sociable', describing the animals' interest in their environment and their engagement with people and pen-mates. PC2 ranged from 'comfortable/relaxed' to 'anxious/nervous/stressed', a demeanour indicating a state of comfort *versus* anxiety of dogs in the kennel environment. PC3 describes fearful demeanour summarised by the terms 'fearful/hesitant/wary', whilst PC4 is characterised by the terms 'depressed/bored', which indicates a negative mood of lower energy than fear, anxious and stress. Reference to emotional states with positive and negative valence has previously been made in assessments of shelter dog welfare: Titulaer et al. [39] for example used cognitive measures to assess the effect of long versus short term

Table 7. Kendall's W coefficients of concordance (n = 11) for the 20 QBA terms.

Variable	Kendall W	Variable	Kendall W
Aggressive	0.489	Fearful	0.561
Alert	0.579	Hesitant	0.524
Anxious	0.686 <sup>a</sup>	Interested	0.503
Attention-seeking	0.679 <sup>a</sup>	Nervous	0.683 <sup>a</sup>
Bored	0.540	Playful	0.563
Comfortable	0.700 <sup>a</sup>	Reactive	0.611 <sup>a</sup>
Curious	0.576	Relaxed	0.689 <sup>a</sup>
Depressed	0.320	Sociable	0.793 <sup>a</sup>
Excited	0.497	Stressed	0.579
Explorative	0.466	Wary	0.714 <sup>a</sup>

<sup>a</sup> Values larger than 0.60

<https://doi.org/10.1371/journal.pone.0212652.t007>

permanence in kennel, whereas Mendl et al. [40] demonstrated how shelter dogs suffering from separation anxiety were more likely to be experiencing a negative affective state. As discussed in the introduction, Walker et al. [27] used comparable QBA dimensions to describe the emotional state of dogs in kennel versus home environment and compare short versus long term housing. Nevertheless, most of the scientific literature on shelter dogs' welfare has so far mainly relied on quantitative ethogram-based behavioural measures (e.g. time spent walking, sleeping etc.) and physiological indicators [41–43].

These four dimensions of dog emotional expressivity address dynamic aspects of welfare including important subtle differentiations, such as that between relaxation and depression, or between emotionally positive and negative excitement (excited vs nervous). From a 'whole-animal' perspective, the aim of QBA is not to identify a minimal set of core descriptive terms, but to capture wider patterns of expression and their context through a larger range of terms. Three terms (i.e. reactive, alert and aggressive) did not load highly on any of the four extracted PCs. Looking at Figs 1 and 2 it appears these terms co-load with stress and anxiety on PC2, and group together on PC4 opposite to 'bored/depressed', thus appearing to mostly reflect a tense reactivity associated with a negative emotional state in dogs struggling to cope with the environment. However, terms such as 'alert' and 'reactive' do not in themselves have a strong negative connotation—animals in positive playful mood can also be alert and reactive, or even mildly aggressive, and this likely explains why these terms often do not necessarily load highly on either positive or negative ends of emotional dimensions. This, however, does not make them superfluous, they can still serve to support and specify patterns of positive and negative mood in dogs assessed in different situations. Overall, the four PCs identified by this study cover the four quadrants of emotional expression defined by valence and arousal axes which tend to be typical of dimensional models of affect [44], and as such should be expected to offer a comprehensive assessment tool of dog emotional expression.

An important element of any measurement tool is its reliability, i.e. the extent to which measures provide consistent results when applied by different assessors [45,46]. To reflect QBA's whole-animal character, reliability is primarily investigated at Principle Component (PC) level, using the PC scores generated through PCA, rather than scores on separate QBA terms. Each of the 4 PCs provides scores for the 13 video clips for each of the 11 observers, and so these scores could be compared. As described in Methods (line 246, this was done in two ways: 1. comparing the ranking of the 13 clips on each PC, i.e. how clip scores were ordered from low to high on each PC, and 2. comparing the quantitative range occupied by clip scores

on each PC, and the mean value resulting from that range. In the present study, agreement on ranking the 13 clips was good for all 4 PCs, ( $W = 0.60\text{--}0.80$ ), which concurs with previous studies testing the reliability of QBA fixed lists in a range of species [25,47,48], though some recent studies indicate that in field conditions as opposed to video-based conditions, achieving good agreement on ranking of animals on QBA dimensions can be more difficult [48,49].

There was however a significant effect of observer on the mean value of clip scores on all 4 PCs, suggesting consistent differences in how observers quantified the various QBA terms using VAS scales. Post-hoc analysis, however, revealed that the majority of observers (60–70%) did not differ significantly in mean clip score on the 4 PCs, with fewer than 4 observers diverging significantly from this mean. Some of these outliers tended to be the same, suggesting these observers had generally different scoring habits across all terms. For PC2 mean scores were quite homogeneous, in fact, there was only one outlier. When running the analysis without that observer, no significant difference between the remaining observers was found for PC2. The higher homogeneity between scores associated with PC2 could be due to PC2 being the only dimension for which high loading terms described a clear contrast between positive and negative expressions. For the other three dimensions the emphasis was on either a positive or a negative expression, providing less conceptual anchoring for observers to place their scores on the visual analogue scale, making variation between observer scores more likely. Other studies have reported variation in 'scoring styles' between observers (e.g. [45]), creating observer effects on mean animal scores on some PCs.

As noted above, observer effects at PC level will mostly be due to differences in how observers use VAS scales to quantify animal expressions on separate QBA terms, and it is therefore useful to also inspect reliability at the level of separate terms. Most QBA studies report low agreement for at least a few terms, even when overall agreement at PC level is good (e.g. [25]). In the present study, 12 out of 20 individual terms showed moderate agreement between 0.50 and 0.60, while 3 terms (depressed, explorative, aggressive) fell below 0.50. The lack in agreement for these 3 terms may be due to the low prevalence of these expressions in the video clips, or, more generally, may indicate that observers found these expressions difficult to assess. Recognising aggression for example has also been difficult for observers in other studies of dog expression [30,50]. However low initial reliability does not necessarily imply a term should be removed from the list, as it may reflect an important aspect of an animal's experience and well-being. Instead efforts could be made to improve a term's descriptive characterisation and provide better training materials.

Generally, the issue of validity, i.e. the question to which extent observers are capable of correctly identifying dog emotional expressions, remains open to debate [29,46]. Some studies found most observers to be reasonably skilled while others did not (e.g. [30,50,51]). One relevant factor might be the method in which observers are presented with information. Some studies instruct observers to focus on particular details of a dog's expression (e.g. facial expression, play), and may present these in relatively fixed format (e.g. photographs or verbal descriptions), whereas other studies ask observers to rate the whole moving animal in different contexts shown on video [45]. QBA falls in the latter category, hypothesizing that observing the dynamic whole animal in context provides more expressive clues for interpretation than focusing on isolated behaviours and signals (e.g. [35]). The current study supports the viability of this approach, but also indicates the need for attention to specific terms and aspects of expression that may detract from the overall reliability of the QBA tool.

This is particularly true when scoring animals live, in the far more variable conditions of practical applied settings, where observers are more likely to be distracted and influenced by different aspects of the physical and social environment [48,49]. However, research on QBA assessment in donkeys and goats under field conditions has shown that training observers, i.e.

taking time to discuss the meaning of individual terms and to compare how these terms are used for scoring, *while located in the field conditions*, can significantly improve the reliability of such terms [25]. Furthermore, as dogs are known to respond differently to male and female observers [52], it is important to adjust any live training exercises for such differences, so that an observer's gender does not unduly bias a dog's assessment. Thus, it is clear that any future practical application of the QBA term list for shelter dog welfare proposed in this study should provide ample training (including field assessments), until consistently high agreement between observers can be reached.

In recent efforts to promote the experience of positive emotions in captive/domestic animals (e.g. through appropriate environment enrichment), welfare scientists have directed a lot of attention toward the validation of outcome (i.e. animal-based) measures that could be integrated into welfare assessment tools to ensure that standards of positive welfare are maximised [53]. QBA has been successfully applied to farm animal welfare assessment protocols [22] and it has the potential to provide a valuable measure of dogs' behavioural expression in confinement. Effective interventions to minimise poor welfare in shelter dogs should be based on the evaluation of dogs' individual ability to cope and adapt to confinement in kennels. QBA focuses on the dynamic expressivity of behavioural demeanour, describing and quantifying the emotional connotation of this expressivity that we naturally perceive. The comprehensive range of qualitative terms developed in this study may function as an integrative screening tool of kennelled dogs' welfare, with special attention to positive welfare. A low expression of certain terms might prompt the implementation of certain types of enrichment that may have been neglected. For example, it has been demonstrated that stress upon entering a rehoming shelter can be mitigated by short sessions of positive human interaction [5]. Hence, low overall scores on the terms characterising PC1 may raise awareness for the need of targeted socialisation programs on animals struggling to adapt to the novel kennel environment. The QBA pre-fixed list developed in this study could be used as a monitoring system to detect early warning signs of reduced welfare state in kennelled dogs, as well as detect areas of improvement to ensure good quality of life.

## Conclusions

The present study aligns with previous QBA studies on other animal species in finding mainly good inter-observer reliability for a fixed-list of QBA terms applied to video-based assessments of dogs living in a shelter environment. Agreement in ranking dogs on the four expressive dimensions was good, but a few observers produced significantly different mean scores for observed dogs on the four main PCs, indicating a need for training and alignment of observer 'scoring styles' [47]. The scientific community recognises that welfare is a complex multidimensional concept, and that no single indicator can be considered exhaustive to evaluate the welfare of animals. It will be preferable therefore to apply QBA in conjunction with other physiological, behavioural or health indicators for animal welfare [17]. The QBA scoring tool developed and tested in the present study can potentially be integrated into existing welfare assessment protocols for shelter dogs, and strengthen the power of those protocols to assess and evaluate the animals' experience [7]. However, to ensure acceptable inter-observer reliability of these protocols, on-going training in the practical application of the QBA tool should be provided until consistently high agreement between observers can be reached. Further research should clarify whether and how reliable use of this tool for can be achieved.

## Acknowledgments

The authors wish to thank the group of experts that anonymously participated in this study as well as all the observers for their contribution.



## Author Contributions

**Conceptualization:** Laura Arena, Shanis Barnard.

**Data curation:** Laura Arena, Françoise Wemelsfelder, Shanis Barnard.

**Formal analysis:** Laura Arena, Françoise Wemelsfelder, Shanis Barnard.

**Investigation:** Laura Arena, Stefano Messori.

**Methodology:** Laura Arena, Stefano Messori, Shanis Barnard.

**Project administration:** Nicola Ferri.

**Supervision:** Françoise Wemelsfelder, Nicola Ferri, Shanis Barnard.

**Validation:** Laura Arena, Françoise Wemelsfelder, Stefano Messori, Nicola Ferri, Shanis Barnard.

**Writing – original draft:** Laura Arena.

**Writing – review & editing:** Françoise Wemelsfelder, Nicola Ferri, Shanis Barnard.

## References

1. Taylor KD, Mills DS. The effect of the kennel environment on canine welfare: A critical review of experimental studies. *Anim Welf.* 2007; 16: 435–447.
2. Bollen KS, Horowitz J. Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. *Appl Anim Behav Sci.* Elsevier; 2008; 112: 120–135. <https://doi.org/10.1016/J.APPLANIM.2007.07.007>
3. Kiddie JL, Collins LM. Development and validation of a quality of life assessment tool for use in kennelled dogs (*Canis familiaris*). *Appl Anim Behav Sci.* 2014; 158: 57–68. <https://doi.org/10.1016/j.applanim.2014.05.008>
4. Kiddie J, Collins L. Identifying environmental and management factors that may be associated with the quality of life of kennelled dogs (*Canis familiaris*). *Appl Anim Behav Sci.* Elsevier B.V.; 2015; 167: 43–55. <https://doi.org/10.1016/j.applanim.2015.03.007>
5. Coppola CL, Grandin T, Enns RM. Human interaction and cortisol: Can human contact reduce stress for shelter dogs? *Physiol Behav.* 2006; 87: 537–541. <https://doi.org/10.1016/j.physbeh.2005.12.001> PMID: 16430930
6. Hewson CJ, Hiby EF, Bradshaw JWS. Assessing quality of life in companion and kennelled dogs: a critical review. *Anim Welf.* 2007; 16: 89–95.
7. Barnard S, Pedernera C, Candeloro L, Ferri N, Velarde A, Villa PD. Development of a new welfare assessment protocol for practical application in long-term dog shelters. *Vet Rec.* 2016; 178: 18. <https://doi.org/10.1136/vr.103336> PMID: 26612859
8. Broom DM. The scientific assessment of animal welfare. *Appl Anim Behav Sci.* Elsevier; 1988; 20: 5–19.
9. Duncan IJH. The changing concept of animal sentience. *Appl Anim Behav Sci.* Elsevier; 2006; 100: 11–19. <https://doi.org/10.1016/J.APPLANIM.2006.04.011>
10. Dawkins MS. Animal Minds and Animal Emotions. *Am Zool.* 2000; 40: 883–888.
11. Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, et al. Assessment of positive emotions in animals to improve their welfare. *Physiol Behav.* 2007; 92: 375–397. <https://doi.org/10.1016/j.physbeh.2007.02.003> PMID: 17428510
12. Yeates JW, Main DCJ. Assessment of positive welfare: a review. *Vet J.* Elsevier; 2008; 175: 293–300. <https://doi.org/10.1016/j.tvjl.2007.05.009> PMID: 17613265
13. Napolitano F, Knierim U, Grasso F, De Rosa G. Positive indicators of cattle welfare and their applicability to on-farm protocols; *Italian Journal of Animal Science* 2009; 8: 355–365.
14. Meagher RK. Observer ratings: Validity and value as a tool for animal welfare research. *Appl Anim Behav Sci.* Elsevier; 2009; 119: 1–14. <https://doi.org/10.1016/J.APPLANIM.2009.02.026>
15. Wemelsfelder F, Hunter TEA, Mendl MT, Lawrence AB. Assessing the ‘whole animal’: a free choice profiling approach. *Anim Behav.* Academic Press; 2001; 62: 209–220. <https://doi.org/10.1006/ANBE.2001.1741>

16. Phythian CJ, Michalopoulou E, Duncan JS, Wemelsfelder F. Inter-observer reliability of Qualitative Behavioural Assessments of sheep. *Applied Animal Behaviour Science* 2013; 144: 73–79.
17. Wemelsfelder F, Mullan S. Applying Ethological and Health Indicators to Practical Animal Welfare Assessment. *Sci Tech Rev Off Int des Epizoot.* 2014; 33: 111–120.
18. Fleming PA, Clarke T, Wickham SL, Stockman CA, Barnes AL, Collins T et al. The contribution of qualitative behavioural assessment to appraisal of livestock welfare. *Animal Production Science*; 2016; 56: 1569–1578. <http://dx.doi.org/10.1071/AN15101>
19. Wemelsfelder F. How Animals Communicate Quality of Life: The Qualitative Assessment of Behaviour. *Anim Welf.* 2007;S: 25–31.
20. Wemelsfelder F, Hunter AE, Paul ES, Lawrence AB. Assessing pig body language: Agreement and consistency between pig farmers, veterinarians, and animal activists1. *J Anim Sci.* Oxford University Press; 2012; 90: 3652–3665. <https://doi.org/10.2527/jas.2011-4691> PMID: 22745187
21. Hintze S, Murphy E, Bachmann I, Wemelsfelder F, Würbel H. Qualitative Behaviour Assessment of horses exposed to short-term emotional treatments. *Appl Anim Behav Sci.* Elsevier; 2017; 196: 44–51. <https://doi.org/10.1016/J.APPLANIM.2017.06.012>
22. Keeling L, Evans A, Forkman B, Kjaernes U. Welfare Quality® principles and criteria. Improving farm animal welfare. Wageningen: Wageningen Academic Publishers; 2013. pp. 91–114. [https://doi.org/10.3920/978-90-8686-770-7\\_5](https://doi.org/10.3920/978-90-8686-770-7_5)
23. Battini M, Vieira A, Barbieri S, Ajuda I, Stilwell G, Mattiello S. Invited review: Animal-based indicators for on-farm welfare assessment for dairy goats. *J Dairy Sci.* Elsevier; 2014; 97: 6625–6648. <https://doi.org/10.3168/jds.2013-7493> PMID: 25242423
24. Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach MC. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. Hillman E, editor. *PLoS One.* Public Library of Science; 2014; 9: e92281. <https://doi.org/10.1371/journal.pone.0092281> PMID: 24647606
25. Minero M, Dalla Costa E, Dai F, Murray LAM, Canali E, Wemelsfelder F. Use of Qualitative Behaviour Assessment as an indicator of welfare in donkeys. *Appl Anim Behav Sci.* Elsevier; 2016; 174: 147–153. <https://doi.org/10.1016/J.APPLANIM.2015.10.010>
26. Walker J, Dale A, Waran N, Clarke N, Farnworth M, Wemelsfelder F. The assessment of emotional expression in dogs using a Free Choice Profiling methodology. *Anim Welf.* 2010; 19: 75–84.
27. Walker JK, Dale AR, D'Eath RB, Wemelsfelder F. Qualitative Behaviour Assessment of dogs in the shelter and home environment and relationship with quantitative behaviour assessment and physiological responses. *Appl Anim Behav Sci.* 2016; 184: 97–108. <https://doi.org/10.1016/j.applanim.2016.08.012>
28. Arena L, Wemelsfelder F, Messori S, Ferri N, Barnard S. Application of Free Choice Profiling to assess the emotional state of dogs housed in shelter environments. *Appl Anim Behav Sci.* 2017; 195: 72–79. <https://doi.org/10.1016/j.applanim.2017.06.005>
29. Konok V, Nagy K, Miklósi Á. How do humans represent the emotions of dogs? The resemblance between the human representation of the canine and the human affective space. *Appl Anim Behav Sci.* Elsevier; 2015; 162: 37–46. <https://doi.org/10.1016/J.APPLANIM.2014.11.003>
30. Tami G, Gallagher A. Description of the behaviour of domestic dog (*Canis familiaris*) by experienced and inexperienced people. *Appl Anim Behav Sci.* <Go to ISI>://WOS:000270421600005; 2009; 120: 159–169.
31. Jones AC, Gosling SD. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Appl Anim Behav Sci.* 2005; 95: 1–53.
32. Kubinyi E, Turcsán B, Miklósi Á. Dog and owner demographic characteristics and dog personality trait associations. *Behav Processes.* 2009; 81: 392–401. <https://doi.org/10.1016/j.beproc.2009.04.004> PMID: 19520239
33. Ley J, Bennett P, Coleman G. Personality dimensions that emerge in companion canines. *Appl Anim Behav Sci.* 2008; 110: 305–317. <https://doi.org/https://doi.org/10.1016/j.applanim.2007.04.016>
34. Ley JM, Bennett PC, Coleman GJ. A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ). *Appl Anim Behav Sci.* Elsevier B.V.; 2009; 116: 220–227. <https://doi.org/10.1016/j.applanim.2008.09.009>
35. Minero M, Tosi MV, Canali E, Wemelsfelder F. Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Appl Anim Behav Sci.* Elsevier; 2009; 116: 74–81. <https://doi.org/10.1016/J.APPLANIM.2008.07.001>
36. Rousing T, Wemelsfelder F. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Appl Anim Behav Sci.* <Go to ISI>://WOS:000242352400004; 2006; 101: 40–53.

37. Minero M., Dalla Costa E., Dai F., Canali E., Barbieri S., Zanella A., et al. Using qualitative behaviour assessment (QBA) to explore the emotional state of horses and its association with human-animal relationship. *Applied Animal Behaviour Science* 204, 53–59.
38. Grosso L, Battini M, Wemelsfelder F, Barbieri S, Minero M, Dalla Costa E, et al. On-farm Qualitative Behaviour Assessment of dairy goats in different housing conditions. *Appl Anim Behav Sci*. Elsevier; 2016; 180: 51–57. <https://doi.org/10.1016/j.applanim.2016.04.013>
39. Titulaer M., Blackwell E.J., Mendl M., Casey R.A., 2013. Cross sectional study comparing behavioural, cognitive and physiological indicators of welfare between short and long term kennelled domestic dogs. *Appl. Anim. Behav. Sci.* 147, 149–158. <https://doi.org/10.1016/j.applanim.2013.05.001>
40. Mendl M, Burman OHP, Paul ES. An integrative and functional framework for the study of animal emotion and mood. *Proc R Soc B Biol Sci.* 2010; 277: 2895–2904. <https://doi.org/10.1098/rspb.2010.0303> PMID: 20685706
41. Bergamasco L, Osella MC, Savarino P, Larosa G, Ozella L, Manassero M, et al. Heart rate variability and saliva cortisol assessment in shelter dog: Human-animal interaction effects. *Appl Anim Behav Sci.* 2010; 125: 56–68. <https://doi.org/10.1016/j.applanim.2010.03.002>
42. Wells D, Hepper PG. The Behaviour of Dogs in a Rescue Shelter. *Anim Welf. Universities Federation for Animal Welfare*; 1992; 1: 171–186.
43. Hiby EF, Rooney NJ, Bradshaw JWS. Behavioural and physiological responses of dogs entering rehoming kennels. *Physiol Behav.* 2006; 89: 385–391. <https://doi.org/10.1016/j.physbeh.2006.07.012> PMID: 16905163
44. Mendl M, Burman OHP, Paul ES. An integrative and functional framework for the study of animal emotion and mood. *Proc R Soc B Biol Sci.* 2010; 277: 2895–2904. <https://doi.org/10.1098/rspb.2010.0303> PMID: 20685706
45. Kaler J, Wassink GJ, Green LE. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. *Vet J. W.B. Saunders*; 2009; 180: 189–194. <https://doi.org/10.1016/J.TVJL.2007.12.028> PMID: 18308594
46. Munch KL, Wapstra E, Thomas S, Fisher M, Sinn DL. What are we measuring? Novices agree amongst themselves (but not always with experts) in their assessment of dog behaviour. *Ethology*; 2019;1–9.
47. Phythian C, Michalopoulou E, Duncan J, Wemelsfelder F. Inter-observer reliability of Qualitative Behavioural Assessments of sheep. *Appl Anim Behav Sci. Elsevier*; 2013; 144: 73–79. <https://doi.org/10.1016/J.APPLANIM.2012.11.011>
48. Czycholl I, Beilage EG, Henning C, Krieter J. Reliability of the qualitative behavior assessment as included in the Welfare Quality Assessment protocol for growing pigs1. *J Anim Sci. Oxford University Press*; 2017; 95: 3445–3454. <https://doi.org/10.2527/jas.2017.1525> PMID: 28805930
49. Muri K, Stubbsjoen SM. Inter-observer reliability of Qualitative Behavioural Assessments (QBA) of housed sheep in Norway using fixed lists of descriptors. *Anim Welf. Universities Federation for Animal Welfare*; 2017; 26: 427–435
50. Bloom T, Friedman H. Classifying dogs' (Canis familiaris) facial expressions from photographs. *Behav Processes*, 2013; 96: 1–10. <https://doi.org/10.1016/j.beproc.2013.02.010> PMID: 23485925
51. Demirbas YS, Ozturk H, Emre B, Kockaya M, Ozvardar T, Scott A. Adults' ability to interpret canine body language during a dog–child interaction. *Anthrozoös*; 2016; 9 581–596, <https://doi.org/10.1080/08927936.2016.1228750>
52. Wells D.L., Hepper P.G., 1999. Male and female dogs respond differently to men and women. *Appl. Anim. Behav. Sci.* 61, 341–349. [https://doi.org/10.1016/S0168-1591\(98\)00202-0](https://doi.org/10.1016/S0168-1591(98)00202-0)
53. Boissy A, Lee C. How assessing relationships between emotions and cognition can improve farm animal welfare. *Rev Sci Tech Int Des Epizoot.* <Go to ISI>://WOS:000337765700010; 2014; 33: 103–110.